



dr hab. inż. Artur Janicki, prof. uczelni  
Instytut Telekomunikacji  
Wydział Elektroniki i Technik Informatycznych  
Politechnika Warszawska  
e-mail: Artur.Janicki@pw.edu.pl  
tel.: +48 22 2347722

Warszawa, 30 maja 2023 r.

## RECENZJA ROZPRAWY DOKTORSKIEJ

mgr. inż. ŁUKASZA PAWLIKA

pt. „*Metody zwiększania skuteczności wirtualnych konsultantów poprzez minimalizację niepoprawnie rozpoznawanych intencji z wypowiedzi klientów*”, przedłożonej Radzie Dyscypliny  
Informatyka Techniczna i Telekomunikacja  
Politechniki Śląskiej

### Przedmiot rozprawy, główne tezy pracy

Praca Pana **mgr. inż. Łukasza Pawlika** pt. „*Metody zwiększania skuteczności wirtualnych konsultantów poprzez minimalizację niepoprawnie rozpoznawanych intencji z wypowiedzi klientów*” dotyczy zagadnienia poprawy jakości systemów dialogowych działających w wirtualnych centrach obsługi (ang. *Contact Center, CC*) obsługujących zapytania pochodzące od klientów np. operatora telekomunikacyjnego.

Przedstawiona rozprawa jest pracą badawczą, z wyraźnym komponentem wdrożeniowym. W swojej rozprawie Autor postawił tezę, że możliwa jest poprawa skuteczności wirtualnych konsultantów w centrum obsługi (CC) poprzez:

1. minimalizację niepoprawnie rozpoznawanych wypowiedzi klientów z zastosowaniem preprocessingu i postprocessingu danych pochodzących z systemu automatycznego rozpoznawania mowy (ang. *automatic speech recognition, ASR*),
2. rozpoznawanie ukrytych intencji klientów na podstawie emocji zawartych w ich wypowiedziach.

Rozprawa została napisana w języku polskim, zawiera 98 stron, 8 rozdziałów, 24 rysunki i 34 tabele.

Rozdział 1. wprowadza w tematykę rozprawy i przedstawia motywację prowadzonych badań.

W rozdziale 2. Doktorant omawia aktualne rozwiązania stosowane w wirtualnych konsultantach w centrach obsługi.

W rozdziale 3. Doktorant definiuje główne cele i tezy rozprawy.

W rozdziale 4. Autor przedstawia formalny model wirtualnego konsultanta.

W rozdziale 5. Doktorant opisuje zaproponowaną metodę poprawy jakości automatycznej transkrypcji mowy w celu poprawy dokładności rozpoznawania intencji.

W rozdziale 6. Doktorant prezentuje zaproponowaną metodę wykrywania ukrytych intencji, wraz z przeprowadzonymi eksperymentami.

W rozdziale 7. Autor przedstawia szczegóły implementacji systemu dialogowego o zwiększonej dokładności rozpoznawanych intencji, stosującego zaproponowane wcześniej metody. Przedstawia szczegółowo wyniki eksperymentów i omawia je.

W rozdziale 8. Autor przedstawia skrótowo podsumowanie osiągnięć oraz wyciąga wnioski.

## **Mocne strony rozprawy**

Wielką zaletą rozprawy jest jej mocny aspekt aplikacyjny. Doktorant jest praktykiem, który od 20 lat jest związany z branżą obsługi klienta w centrach obsługi, dlatego badania były prowadzone w odpowiedzi na bezpośrednie potrzeby przemysłu. Część prac Doktoranta była prowadzona w ramach projektu „EMOTICA AI – Inteligentny System Contact Center” (nr umowy: POIR.04.01.04-00-0079/19-00, finansowanego z konkursu NCBiR w ramach POIR). Z dużego doświadczenia Doktoranta wynika również to, że w pracy znajdujemy sporo praktycznych informacji pod kątem wdrożenia poszczególnych komponentów, np. informacje o rodzajach licencji różnych modułów ASR.

Znaczącym osiągnięciem Doktoranta jest utworzenie i opracowanie (w tym zanonimizowanie) zbioru 1054 rzeczywistych rozmów telefonicznych z ponad 100 godz. nagrań. Byłoby bardzo pożyteczne, gdyby udało się udostępnić ten zbiór do celów badawczych (nie ma w pracy informacji na ten temat).

Potwierdzeniem docenienia jakości osiągnięć Doktoranta jest jego dorobek publikacyjny. Autor opublikował trzy artykuły, wszystkie w czasopismach z wyznaczonym współczynnikiem wpływu (IF), w tym znaczący artykuł „*A method for improving bot effectiveness by recognising implicit customer intent in contact centre conversations*”, którego Doktorant jest pierwszym autorem, opublikowany w czasopiśmie *Speech Communications* w 2022 r. Doktorant nie przywołuje tego artykułu w rozprawie, przypuszczam więc, że ukazał się on już po jej złożeniu.

Oryginalnym rozwiązaniem Doktoranta jest zaproponowana metoda poprawy rozpoznawania intencji klienta wirtualnego centrum obsługi, a w szczególności zaproponowany algorytm rozpoznawania ukrytych intencji, wykorzystujący rozpoznawanie stanu emocjonalnego mówcy.

Uważam, że rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego. Osiągnięte wyniki mogą zostać praktycznie wykorzystane w automatycznych systemach dialogowych, w szczególności przeznaczonych dla centrów obsługi, które stają się coraz

bardziej popularne. Doktorant wykazał również znajomość teoretyczną z dziedziny informatyka techniczna i telekomunikacja.

## Słabe strony rozprawy

Do słabych stron rozprawy zaliczam niewystarczające, według mnie, odniesienie metod opracowanych przez Doktoranta do technik stosowanych przez innych badaczy. Doktorant tłumaczy, że „w trakcie przeglądu literatury nie znaleziono prac badawczych dotyczących poprawy jakości transkrypcji automatycznych dedykowanych dla systemów CC opartych na modelu języka polskiego”. Rozumiem pewną specyfikę wypowiedzi klientów podczas rozmowy z CC po polsku, jednak wg mnie z powodzeniem można by się odnieść do badań dotyczących rozpoznawania mowy z sygnału telefonicznego (np. V. Chunwijitra and N. Kurpukdee, "Improving automatic transcription of call center speech using data simulation," ECTI-CON 2021, Chiang Mai, Thailand, 2021, pp. 842-845, doi: 10.1109/ECTI-CON51831.2021.9454803), prac dotyczących odwrotnej normalizacji tekstu (np. M. Sunkara, C. Shivade, S. Bodapati and K. Kirchhoff, "Neural Inverse Text Normalization," ICASSP 2021, Toronto, Canada, 2021, pp. 7573-7577, doi: 10.1109/ICASSP39728.2021.9414912), czy zagadnień dot. poprawy jakości rozpoznawania mowy w języku polskim (np. T. Ziętkiewicz, "Post-editing and rescoring of ASR results with edit operations tagging". PolEval 2020 Workshop, pp. 23-31, i inne prace w ramach konkursu PolEval).

W innym miejscu rozprawy (str. 19) „Autor podkreśla, że w kontekście problematyki CC nie są mu znane badania, które uwzględniałyby wykorzystanie emocji do wykrywania rzeczywistych intencji z wypowiedzi klienta”. Wg mnie w tym miejscu warto byłoby przyjrzeć się bliżej badaniom dot. rozpoznawania sarkazmu. Google Scholar na zapytanie „sarcasm detection” zwraca 46 tysięcy pozycji (jedną zresztą Doktorant cytuje). Przegląd badań dotyczących rozpoznawania sarkazmu zawiera np. publikacja: P. Verma, N. Shukla, & A. P. Shukla, „Techniques of sarcasm detection: A review”. ICACITE 2021, pp. 968-972.

W rozprawie systemy dialogowe używane w eksperymentach (RASA i DialogFlow) są przedstawione jako czarne skrzynki: nie mamy informacji nt. ich konfiguracji, w szczególności nie wiemy, czy z systemu RASA wykorzystany jest tylko moduł rozpoznawania intencji (oparty na algorytmie DIET), czy również moduł wyboru następnej akcji (wykorzystujący np. algorytm TED). Nie wiemy, jakie były ustawione hiperparametry tego algorytmu, w szczególności długość historii. Algorytm TED z odpowiednio długą historią mógłby być z powodzeniem użyty do porównania z zaproponowanym przez Doktoranta algorytmem rozpoznawania ukrytych intencji. Tymczasem, niestety, Doktorant w rozprawie porównuje wyniki swojej metody tylko w odniesieniu do własnych badań.

Podobnie nie mamy informacji nt. sposobu rozpoznawania emocji. Rys. 7.1 sugeruje, że emocje były rozpoznawane na podstawie zarówno głosu, jak i tekstu, natomiast nie znalazłem żadnych szczegółów dot. tego procesu, np. jakie algorytmy zostały użyte? Jaka była dokładność rozpoznawania emocji?

W swojej rozprawie Doktorant przedstawia również kontrowersyjne tezy (np. że „ludzie są bardziej szczerzy w stosunku do botów głosowych i tekstowych niż w stosunku do realnej osoby”, str. 8) bez powoływania się na źródła. Jak się domyślam, są to jego przypuszczenia oparte na własnych obserwacjach, nie zostały one jednak udowodnione.

Moją wątpliwość budzi także zaproponowane przez Doktoranta przetwarzanie wstępne poprzez usuwanie interpunkcji, zwłaszcza w połączeniu z lematyzacją. W przypadku wypowiedzi „Nie, chcę nową umowę” zostanie ona zapewne przetworzona do „Nie, chcę nowa umowa” – obawiam się, że wtedy intencja klienta nie zostanie prawidłowo rozpoznana.

Jako niedociągnięcie uważam też fakt, że Doktorant stosował miejscami terminologię odmienną od ogólnie przyjętej np. w tematyce uczenia maszynowego: zbiór nagrań służący do optymalizacji swojego rozwiązania nazwał badawczym (RLPHDB-RSR), podczas gdy zbiór służący do testowania rozwiązania nazwał walidacyjnym (RLPHDB-VLD). Tymczasem zbiory te nazywa się odpowiednio walidacyjnym i testowym, więc może to wprowadzać czytelnika w błąd.

Doktorant słusznie wskazuje, że metryka wyrazowej stopy błędów (ang. *word error rate*, WER) jest bardziej miarodajna przy ocenie poprawności działania systemów ASR. Niejasne jest więc, dlaczego dodatkowo używa również metryki literowej stopy błędów (ang. *letter error rate*, LER), która, szczególnie dla języka polskiego, wydaje się mniej przydatna.

W rozprawie występują również nieprecyzyjne sformułowania, np.:

- „nieograniczona natura i wewnętrzna zmienność języka mówionego” (str. 11);
- „skuteczność rozpoznawania mowy została oceniona na 86%” (str. 11) – czy chodziło o dokładność rozpoznawania? Czy skuteczność była liczona w jakiś inny sposób?
- „sygnał taki cechuje zwykle częstotliwość próbkowania na poziomie 8 kHz” – chodzi zapewne o to, że sygnał mowy został spróbkowany z częstotliwością 8 kHz;
- „często wykorzystywano (...) możliwości powiększania widma” (str. 29) – jest to niejasne;
- „liczba zamiany liter” – chodzi pewnie o liczbę zamienionych liter? (str. 31).

Doktorant w rozprawie odwołuje się również do artykułów, które... zapewne nie powstały. Są to pozycje: [82] “Artykuł o rozpoznawaniu emocji (do opublikowania)”, [97] “Artykuł o wyborze konkretnych emocji z psychologami i lingwistą (do opublikowania)”.

Inne, drobniejsze uwagi:

- W pracy występują pewne niedociągnięcia formy – o czym więcej w następnym rozdziale;
- W tabeli 5.1 brakuje wartości bezwzględnych WER. Nie mam też pewności, czy poprawa WER wyrażona jest w procentach czy w punktach procentowych. Przydatne byłoby też pokazanie zmian liczby błędów w poszczególnych kategoriach (usunięcia, zamiany, wstawienia);
- W tabeli 5.9 poprawa, wg mnie, wyrażona jest jednak w punktach procentowych, a nie w procentach, jak zaznaczono w tabeli. Wyniki w procentach byłyby wyższe.

Stwierdzam, że przedstawiona rozprawa doktorska **spełnia** warunki określone w art. 13 ust. 1 Ustawy z dn. 14.03.2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65 r., poz. 595 z późn. zm.), dlatego niniejszym wnioskuję o dopuszczenie Doktoranta, Pana **mgr. inż. Łukasza Pawlika**, do publicznej obrony jego rozprawy doktorskiej.

- Nie wiadomo, według jakiego klucza przedstawiono macierze pomyłek dla wyników z tabeli 5.9. Jeśli wg największej poprawy, to w grupie 1 powinien zostać wybrany system Dialogflow.
- W literaturze częściej stosowany jest termin *character error rate* (CER) zamiast *letter error rate* (LER);
- Przy interpretacji wyników WER Doktorant powołuje się na pozycję literatury [96], która kieruje na strony firmy Microsoft oraz [88], kierującą do Wikipedii. Wskazane byłyby bardziej naukowe źródła (np. książka, czasopismo, artykuł konferencyjny);
- „PSL” (str. 44) to nie skrót, tylko skrótowiec (akronim);
- Wg mnie nie było potrzeby umieszczania analizy wszystkich fraz testowych w treści rozprawy. Ewentualnie można było umieścić je w załączniku.

## Strona edycyjna pracy

Rozprawa doktorska mgr. inż. Łukasza Pawlika jest napisana dość starannie, jej format jest zgodny z ogólnie przyjętymi standardami dla prac naukowych w dziedzinie nauk technicznych. Układ pracy jest logiczny i przyjazny dla czytelnika. Rysunki i tabele są staranne.

W pracy występują jednak pewne niedociągnięcia natury edycyjnej i językowej. Są to m.in.:

- Nieużywanie polskich odpowiedników terminów, które są dobrze ugruntowane w naszym środowisku naukowo-technicznym, takich jak: *wydźwięk* (zamiast „*sentymnt*”, który jest kalką z języka angielskiego), *przetwarzanie wstępne* (zamiast „*preprocessing*”), *macierz pomyłek* (zamiast „*macierz konfuzji*”);
- Błędy językowe, np. „*oparty o sztuczną inteligencję*” zamiast prawidłowej formy: „*oparty na sztucznej inteligencji*”;
- Błędy literowe, np. „*Diagram kolejność*” => „*Diagram kolejności*” (str. 38);
- Naprzemienne używanie terminu *Call Center* i skrótowca *CC*;
- Emocje (radość, smutek, ...) niepotrzebnie są wg mnie pisane wielkimi literami (np. str. 18);
- Błędy interpunkcyjne, np. braki lub nadmiar przecinków;
- Stosowanie karkołomnych akronimów, np. RLPHDB-VLD;
- Męczący dla czytelnika jest fakt, że jest on odsyłany do rysunku, który jest np. 6 stron dalej (np. str. 34, odsyłacz do rys. 5.9).

Uchybienia edycyjne nie umniejszają co prawda dorobku Doktoranta, ale miejscami utrudniają lekturę rozprawy.

## Wnioski końcowe

W podsumowaniu stwierdzam, że problemy badawcze postawione w rozprawie zostały opracowane i w znacznym stopniu rozwiązane, a postawiona w pracy teza została udowodniona. Doktorant zaproponował rozwiązania podwyższające dokładność rozpoznawania intencji użytkownika w systemie dialogowym dla języka polskiego przeznaczonym dla centrum obsługi i udowodnił ich skuteczność.

- Nie wiadomo, według jakiego klucza przedstawiono macierze pomyłek dla wyników z tabeli 5.9. Jeśli wg największej poprawy, to w grupie 1 powinien zostać wybrany system Dialogflow.
- W literaturze częściej stosowany jest termin *character error rate* (CER) zamiast *letter error rate* (LER);
- Przy interpretacji wyników WER Doktorant powołuje się na pozycję literatury [96], która kieruje na strony firmy Microsoft oraz [88], kierującą do Wikipedii. Wskazane byłyby bardziej naukowe źródła (np. książka, czasopismo, artykuł konferencyjny);
- „PSL” (str. 44) to nie skrót, tylko skrótowiec (akronim);
- Wg mnie nie było potrzeby umieszczania analizy wszystkich fraz testowych w treści rozprawy. Ewentualnie można było umieścić je w załączniku.

## Strona edycyjna pracy

Rozprawa doktorska mgr. inż. Łukasza Pawlika jest napisana dość starannie, jej format jest zgodny z ogólnie przyjętymi standardami dla prac naukowych w dziedzinie nauk technicznych. Układ pracy jest logiczny i przyjazny dla czytelnika. Rysunki i tabele są staranne.

W pracy występują jednak pewne niedociągnięcia natury edycyjnej i językowej. Są to m.in.:

- Nieużywanie polskich odpowiedników terminów, które są dobrze ugruntowane w naszym środowisku naukowo-technicznym, takich jak: *wydźwięk* (zamiast „*sentymnt*”, który jest kalką z języka angielskiego), *przetwarzanie wstępne* (zamiast „*preprocessing*”), *macierz pomyłek* (zamiast „*macierz konfuzji*”);
- Błędy językowe, np. „*oparty o sztuczną inteligencję*” zamiast prawidłowej formy: „*oparty na sztucznej inteligencji*”;
- Błędy literowe, np. „*Diagram kolejność*” => „*Diagram kolejności*” (str. 38);
- Naprzemienne używanie terminu *Call Center* i skrótowca *CC*;
- Emocje (radość, smutek, ...) niepotrzebnie są wg mnie pisane wielkimi literami (np. str. 18);
- Błędy interpunkcyjne, np. braki lub nadmiar przecinków;
- Stosowanie karkołomnych akronimów, np. RLPHDB-VLD;
- Męczący dla czytelnika jest fakt, że jest on odsyłany do rysunku, który jest np. 6 stron dalej (np. str. 34, odsyłacz do rys. 5.9).

Uchybienia edycyjne nie umniejszają co prawda dorobku Doktoranta, ale miejscami utrudniają lekturę rozprawy.

## Wnioski końcowe

W podsumowaniu stwierdzam, że problemy badawcze postawione w rozprawie zostały opracowane i w znacznym stopniu rozwiązane, a postawiona w pracy teza została udowodniona. Doktorant zaproponował rozwiązania podwyższające dokładność rozpoznawania intencji użytkownika w systemie dialogowym dla języka polskiego przeznaczonym dla centrum obsługi i udowodnił ich skuteczność.

Stwierdzam, że przedstawiona rozprawa doktorska **spełnia** warunki określone w art. 13 ust. 1 Ustawy z dn. 14.03.2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65 r., poz. 595 z późn. zm.), dlatego niniejszym wnioskuje o dopuszczenie Doktoranta, Pana **mgr. inż. Łukasza Pawlika**, do publicznej obrony jego rozprawy doktorskiej.